

# 13

## Essential Elements and Ethical Principles for Trustworthy Artificial Intelligence Adoption in Courts

---

**Carlos E. Jimenez-Gomez**

Judicial Sector and Information Systems Expert,  
National Center for State Courts

**Jesus Cano Carrillo**

Chief Information Officer,  
Constitutional Court of Spain



**Artificial Intelligence-based tools are already being used by courts in the United States. The importance and challenges of these technologies, including legal and ethical ones, require special attention and urgent steps.**

Tasks in courts have rapidly evolved from manual to digital work. In these innovation processes, theory and practice have demonstrated that adopting technology *per se* is not the right path. Innovation in courts requires specific plans for digital transformation, including analysis, programmatic changes, or skills. Artificial Intelligence (AI) is not an exception.

The use of AI in courts is not futuristic. From efficiency to decision-making support, AI-based tools are already being used by U.S. courts. To cite some examples, AI tools allow the discovery of divergences, disparities, and dissonances in jurisdictional activity. At a higher level, AI helps improve internal organization. AI helps with judicial decision consistency, exploiting a large judicial knowledge base in the form of big data, and it makes the judge's work more agile with pattern and linguistic recognition in documents, identifying schemes and conceptualizations.

AI could bring considerable benefits to the judicial system. However, the risks and challenges are also enormous, posing unique hurdles for user trust. Some of the most internationally controversial and discussed

cases are from the United States.<sup>1</sup> These tools could even impact important aspects such as the due process of law,<sup>2</sup> including potential discrimination (see U.S. Government Accountability Office, 2021; Chohlas-Wood, 2020). The legal and ethical implications of technology will play an important role to protect citizens' rights. These implications are intimately linked to civil and human rights' guarantees and basic principles of the rule of law.

As the European Commission for the Efficiency of Justice (2018) underlines on the use of AI in judicial systems, it is critical to preserve “the guarantees of the rule of law, together with the quality of public justice.” Therefore, we must be prepared for the proper adoption and use of AI-based technologies, understanding first the inferred meaning, before deciding if, when, and how AI should be used.

This article defines AI in relation to courts to understand challenges and implications and reviews AI components with a special focus on characteristics of trustworthy AI. It also examines the importance of a new policy and regulatory framework, and makes recommendations to avoid major problems.

## What Does Artificial Intelligence Mean, and Why Are Data So Important?

The judicial sector is increasingly aware of the importance of the data-related ecosystem and its applicability, including elements like open data, data management, data governance, or data standardization. It is also relevant to understand AI's role and how it is linked to the courts. The magnifying effect of these technologies can be very positive, but they could generate unexpected negative effects if they are not used correctly.

---

<sup>1</sup> For example, risk assessment instruments, such as Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). See also, Završnik, 2020.

<sup>2</sup> *State v. Loomis* 881 N.W.2d 749 (Wis. 2016); *State of Kansas v. John Keith Walls*, Opinion No. 116,027, Court of Appeals of the State of Kansas (2017).

## ESSENTIAL ELEMENTS AND ETHICAL PRINCIPLES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE ADOPTION IN COURTS

Improvements in algorithms, data science, big data, and distributed data processing have brought AI to reality (Jimenez-Gomez, Cano-Carrillo, and Falcone-Lanas, 2020). But what does AI exactly mean? AI is a concept used for technologies able to autonomously make predictions, explanations, or recommendations. AI is based on data and advanced algorithms. According to Organisation for Economic Co-operation and Development (OECD) (2019), an AI system may be a “machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.”

To use a cooking analogy, the *algorithm* is the recipe, the *data* are the ingredients, and the resulting *final model* is the finished dish. This cooking analogy is only useful to understand the components and the importance of how they are combined. In general, the models can be based on supervised, unsupervised, or reinforcement learning techniques. The process of building a final AI model is recursive and complex. Depending on the learning technique, models can even be capable of learning from their errors.

Different data sets are used first to train the model and then to test and validate the final model. Results are based on statistical analysis and mathematical probabilities, which means that levels of *accuracy* and *errors* are components of the process. Following the culinary example, once a model has been trained, it is possible to estimate (predict, explain, or recommend) whether a particular dish is at its proper cooking point, or how much closer (what percentage) it is to the desired degree of doneness.

AI systems learn from the data selected for training the model. Therefore, data are an essential factor to be considered before an AI system is created. Data sets have a very high level of sensitivity in determining results. It is critical to have guarantees along the full process, from data selection to the results, because a fine line separates a trustworthy result from a dangerously biased result.

Data will serve to train and build the models that will become the tools used to autonomously make predictions, explanations, or recommendations. There is no AI without data. Data sets used to train AI systems are critical to avoid future errors and biases. Data elements such as quality, accuracy, or completeness will play a critical role in developing AI technologies.

In this new paradigm, ethics is an essential principle. The OECD (2021) and the United Kingdom Government Digital Service (2020) explain the importance of ethics, data, and data science as central elements in the design and use of AI tools in public organizations. In the field of justice, this characteristic has also been highlighted by the European Ethical Charter

on the Use of Artificial Intelligence In Judicial Systems and Their Environment (European Commission, 2018). Indeed, the ethical component is so relevant that ethics have already been placed at the core of new AI international standards.<sup>3</sup>

## Key Principles for Trustworthy AI Adoption in Judicial Systems

The nature, implications, and components of AI tools will require observing special guarantees and agreed-upon principles. From interoperability to audit operations, trustworthiness will require specific attributes that will have to be taken into account for AI adoption.<sup>4</sup>

AI tools should reinforce the idea that information and computing enable effective judicial protection, as well as greater access to justice. For example, it could bring us to more informed justice decisions based on accurate data or the improvement of processes at judicial, administrative, and technological levels. A *smart justice* concept should include not only intelligent technologies but also a “social smartness” to protect the rights of and provide the best services to citizens. This component should include legal and social digital services.

To that end, the “European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment” addresses different principles related to AI adoption: compatibility with fundamental rights (including the design and implementation processes); nondiscrimination (i.e., individuals, groups, and sensitive data); quality and security (i.e., traceability sources of information, certified data, and secure environments); transparency, impartiality, and fairness (data-processing methods explainable and auditable); and “under user control” (informed, with control on results).

---

<sup>3</sup> Within the IEEE Ethics in Action in Autonomous and Intelligent Systems, standards: IEEE 7000-2021 IEEE Standard Model Process for Addressing Ethical Concerns during System Design; IEEE 7001 on Transparency of Autonomous Systems; and IEEE P7003 on Algorithmic Bias Considerations.

<sup>4</sup> “Interoperability” is understood as the basis for shared services, information, and data between different sources and public bodies.



## ESSENTIAL ELEMENTS AND ETHICAL PRINCIPLES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE ADOPTION IN COURTS

In the United States, the Institute of Electrical and Electronics Engineers (IEEE-USA, 2020) Artificial Intelligence Policy Committee recommends the development and adoption of explicit risk-benefit analysis frameworks, as well as a set of recommendations, standards, and principles.<sup>5</sup> The National Institute of Standards and Technology is currently developing a framework to manage risks, proposing a methodology to consider an AI task's risk level and the user decision, based on nine factors that contribute to a human's potential trust in an AI system: accuracy, reliability, resiliency, objectivity, security, explainability, safety, privacy, and accountability (Stanton and Jensen, 2021).

Other organizations, like the Government Accountability Office (GAO, 2021) or the [Joint Technology Committee](#) (JTC, 2020), are likewise highlighting the importance of different AI adoption components. Describing practices for federal agencies and other entities, GAO is proposing an Artificial Intelligence Accountability Framework that highlights four complementary components to be addressed: governance through accountability; data in terms of quality, reliability, and representativeness; performance through consistent results; and monitoring for reliability and relevance. JTC talks about *common sense and ethics*, referring mainly to the U.S. Department of Defense (DoD) Ethical Principles for Artificial Intelligence: responsible, equitable, traceable, reliable, and governable. Finally, other authors, like Wing (2021), talk about a broad set of overlapped properties, including accuracy, robustness, fairness, accountability, transparency, interpretability and explainability, and ethics, as well as reliability, safety, security, privacy, availability, and usability.

Principles and components follow different perspectives and granularity levels, depending on the approach (see Table 1). An analysis of the different cases shows links between components, as seen in the IEEE-USA or NIST elements, which could be also explained within the European Commission (2018) ethical charter. Therefore, we should pay attention to a comprehensive perspective, including both explicit and implicit (and underlying) components that are also linked to social and legal implications of these technologies.

---

<sup>5</sup> See also IEEE Ethically Aligned Design initiative.

## ESSENTIAL ELEMENTS AND ETHICAL PRINCIPLES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE ADOPTION IN COURTS

**Table 1: Summary of Components Highlighted**

	European Commission (EU-Justice)	GAO (US Gov- Administration)	IEEE-USA (Standards)	NIST (US Gov- Technology)	US DoD (US Gov-Defense)	Wing, J. (2021)
Accountability			✓	✓		✓
Accuracy				✓		✓
Availability						✓
Awareness of misuse			✓			
Competence			✓			
Data (quality, reliability, representativeness)		✓				
Data agency control			✓			
Effectiveness			✓			
Equitable (minimize bias)					✓	
Ethical						✓
Explainability				✓		
Fairness						✓
Fundamental rights	✓					
Governable (control)					✓	
Governance (accountability)		✓				
Human rights			✓			
Interpretability/Explainability						✓
Monitoring (reliability and relevance)		✓				
Non-discrimination	✓					
Objectivity				✓		
Other						✓
Other: Auditing standards		✓				
Other: Open data	✓					
Other: Openness			✓			
Other: Organizational measures					✓	
Other: Responsible use		✓				
Other: Risk management	✓	✓	✓	✓	✓	
Other: Standards	✓		✓	✓	✓	
Performance (consistent results)		✓				
Privacy				✓		✓
Quality and security	✓					
Reliability				✓		✓
Reliable (safety, security, effectiveness)					✓	
Resiliency				✓		
Responsible (judgment and competence)					✓	
Robustness						✓
Safety				✓		✓
Security				✓		✓
Traceable (transparency and audit)					✓	
Transparency			✓			✓
Transparency, impartiality, and fairness	✓					
Usability						✓
User control	✓					
Well-being			✓			

# The Human Factor, the Organizational Implications, and the Importance of a New Policy and Regulatory Framework

To achieve the goals of the justice system and provide better services to citizens and society, AI adoption will require courts to be aware of a multidisciplinary perspective. The socio-technical implications of technology will play a key role for the rule of law and the quality of the justice in preserving citizens' rights. Critical components are also linked to what could be called the "human factor." Policymakers, designers, and implementers can shape the AI process and results because of their adoption or rejection of the results offered by AI tools.

As Reiling (2020) states, judges need to understand how AI works. Recent research shows the importance of being well prepared for the correct adoption and use of AI. According to Raunak and Kuhn (2021), AI users could either overly trust the results of AI systems and follow incorrect recommendations, or overly distrust them and reject correct recommendations. Green and Chen (2021) highlight how AI-based tools can even alter human decision-making processes in harmful ways.

The need for trustworthy AI is defining the right path to be followed by courts. This complex path will be facilitated by the adoption of a specific new generation of standards, rules, norms, protocols, and mechanisms where human factors and the organizational context will play an important role. These essential elements are closer to social attributes than technological components, moving from skills to governance of organizational-related components.

Innovations in digital justice, new digital procedures, and new digital tools, like videoconferencing or digital signatures, can even require regulatory changes based on the legal implications of technologies. With AI we probably are facing one of the most disruptive and transformative components in the history of justice and courts' modernization. To preserve the rule of law and, eventually, the rights of our citizens in our society, AI adoption in courts will require an urgent review of the regulatory and policy framework, as well as guidelines on the use of these technologies.



## ESSENTIAL ELEMENTS AND ETHICAL PRINCIPLES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE ADOPTION IN COURTS

A new regulation on AI in the European Union mentions the importance of addressing risks of potential biases, errors, and opacity in the administration of justice, highlighting explicitly that “certain AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial.”<sup>6</sup> The regulation seeks to ensure that AI systems respect the existing laws on fundamental rights, enhance governance and effective law enforcement, and facilitate the development of lawful, safe, and trustworthy AI.

---

<sup>6</sup> Regulation of the European Parliament and of the Council. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM (2021) 206 final.



# Recommendations

We have seen that the use and adoption of trustworthy AI tools require a multidimensional approach, including technical, organizational, legal, governance, and strong data and ethical-related components. The ethical perspective is not secondary, and it should be seen as a core element, especially in the courts. Based on the previous points discussed, we conclude with some recommendations:

- ***Develop a trustworthy AI adoption framework, including AI risks management and mitigation incorporating AI guidelines and standards from internationally trusted sources such as IEEE.*** It should guide courts on key principles and elements, considering the specifics of the U.S. judiciary and courts. To solve complex legal problems while avoiding new ones, court IT managers will have to take very seriously the criteria on how these tools are designed, developed, implemented, used, monitored, and evaluated.
- ***Prepare a trustworthy AI regulatory framework, implementing agreed-upon policies and recommendations from sources like CCJ/COSCA and JTC, addressed to reinforce the ethical and trustworthy AI adoption in courts.*** Additional regulation in fields such as privacy or security will play also an important role.
- ***Adopt a strategic digitalization perspective for a data-driven court.*** This includes policies focused on data quality; data management or data governance; new roles such as data scientist or chief data officer; and strong mechanisms of data-driven AI techniques evaluation, for direct support of jurisdictional decision making.
- ***Advance toward an open justice perspective also linked to AI, efficiency, and interoperability.*** Sharing digital services, information, and data will require actions addressed to transparency and accountability in the use of models, algorithms, and data and adoption of open standards on data, taking advantage of initiatives like the NCSC [National Open Court Data Standards](#).
- ***Strengthen AI-related literacy, knowledge, skills, and capabilities in the workforce.*** Understanding these technologies and principles will be critical to use AI adequately.

## References

Chohlas-Wood, A. (2020). "[Understanding Risk Assessment Instruments in Criminal Justice](https://perma.cc/L8KK-84S3)." Report, Brookings Institution, Washington, D.C., June 19. Perma link: <https://perma.cc/L8KK-84S3>.

European Commission for the Efficiency of Justice (2018). "[European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment](https://perma.cc/5VSD-AXMB)." Council of Europe. Perma link: <https://perma.cc/5VSD-AXMB>.

Green, B., and Y. Chen (2021). "[Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts](https://perma.cc/ZGR2-9X3L)." *Proceedings of the ACM on Human-Computer Interactions*, October. Perma link: <https://perma.cc/ZGR2-9X3L>.

Harvard Law Review (2017). "[State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing](https://perma.cc/A3DK-M5JW)." 130 *Harvard Law Review* 1530. Perma link: <https://perma.cc/A3DK-M5JW>.

Institute of Electrical and Electronics Engineers (IEEE-USA, 2020). "[Artificial Intelligence: Accelerating Inclusive Innovation by Building Trust](https://perma.cc/PWD3-ENKF)." Position statement. Adopted by IEEE Board of Directors, July 21. Perma link: <https://perma.cc/PWD3-ENKF>.

Jimenez-Gomez, C.E., J. Cano-Carrillo, and F. Falcone-Lanas (2020). "[Artificial Intelligence in Government](https://perma.cc/4PCH-MHSZ)." *Computer*, October, 23-27.

Joint Technology Committee (JTC, 2020). "[Introduction to AI for Courts](https://perma.cc/4PCH-MHSZ)." *JTC Resource Bulletin*, March 27. Perma link: <https://perma.cc/4PCH-MHSZ>.

Kehl, D., P. Guo, and S. Kessler (2017). "[Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing](https://perma.cc/76CS-ZFTW)." Responsive Communities Initiative, Berkman Klein Center for Internet and Society, Harvard Law School. Perma link: <https://perma.cc/76CS-ZFTW>.

Organisation for Economic Co-operation and Development (2021). [Good Practice Principles for Data Ethics in the Public Sector](https://perma.cc/NB9E-JD8A). Digital Government and Data Unit. Perma link: <https://perma.cc/NB9E-JD8A>.

— (2019). [Recommendation of the Council on Artificial Intelligence](https://perma.cc/76CS-ZFTW). Adopted May 21. OECD/LEGAL/0449.



Reiling, A. D. (2020). "Courts and Artificial Intelligence." 11(2) *International Journal for Court Administration* 8.

Raunak, M., and R. Kuhn (2021). "[Explainable Artificial Intelligence and Machine Learning](#)." *Computer*, October, 25-27.

Stanton, B., and T. Jensen (2021). "[Trust and Artificial Intelligence](#)." Interagency/Internal report (NISTIR). National Institute of Standards and Technology, Gaithersburg, Maryland. Perma link: <https://perma.cc/5PTK-LNF7>.

United Kingdom Government Digital Service (2020). "[Data Ethics Framework](#)." Perma link: <https://perma.cc/FTK7-KLNQ>.

U.S. Government Accountability Office (GAO, 2021). [Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities](#). Washington, DC: GAO. GAO-21-519SP. Perma link: <https://perma.cc/X5A6-K6C6>.

Wing, J. M. (2021). "[Trustworthy AI](#)." *Communications of the ACM*, October, 64-71. Perma link: <https://perma.cc/HXR6-2CQ5>.

Završnik, A. (2020). "Criminal Justice, Artificial Intelligence Systems, and Human Rights." 20 *ERA Forum* 567.